

STAC67: Regression Analysis

Lecture 12

Thibault Randrianarisoa

October 16, 2026

University of Toronto Scarborough

Review of Lecture 11

At a new point $\mathbf{X}_h$, the mean response is estimated by $\hat{Y}_h = \mathbf{X}_h' \hat{\beta}$, with

$$s^2\{\hat{Y}_h\} = MSE \mathbf{X}_h' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_h, \quad s^2\{\text{pred}\} = MSE (1 + \mathbf{X}_h' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_h).$$

Least squares is the **orthogonal projection** of $\mathbf{Y}$ on the column space $C(\mathbf{X})$:

$$\hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y}, \quad \mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y} \perp C(\mathbf{X}), \quad \mathbf{X}'\mathbf{e} = \mathbf{0} \text{ (the normal equations).}$$

$$H^2 = H$$

$\mathbf{H}$ is symmetric and idempotent, with $\mathbf{H}\mathbf{X} = \mathbf{X}$. $\mathbf{I} - \mathbf{H}$ projects on the $(n - p)$ -dimensional space orthogonal to $C(\mathbf{X})$, where $\mathbf{e}$ lives.

Today: we finish Lecture 11 (the residuals, Pythagoras, degrees of freedom), then tests and intervals, and the overall F test.

Residuals and the analysis of variance

Properties of the residuals

$$E[e] = (I - H)E[Y] = (I - H)X\beta = X\beta - HX\beta = 0$$

$e = (I - H)Y$ is again a linear function of Y . Since $(I - H)X = 0$:

$$E(e) = (I - H)X\beta = 0, \quad \sigma^2\{e\} = \sigma^2(I - H),$$

and with normal errors, $e \sim N(0, \sigma^2(I - H))$.

The proof of $\sigma^2\{e\}$ is Assignment 2, Question 3(c).

$$\begin{aligned} \sigma^2\{e\} &= (I - H)(\sigma^2 I)(I - H) \\ &= \sigma^2 (I - H)(I - H) \\ &= \sigma^2 (I - H)^2 \\ &= \sigma^2 (I - H) \end{aligned}$$

(idempotent)

Properties of the residuals

$e = (I - H)Y$ is again a linear function of Y . Since $(I - H)X = 0$:

$$E(e) = (I - H)X\beta = 0, \quad \sigma^2\{e\} = \sigma^2(I - H),$$

and with normal errors, $e \sim N(0, \sigma^2(I - H))$.

The proof of $\sigma^2\{e\}$ is Assignment 2, Question 3(c).

Write h_{ii} for the diagonal entries of H . On the diagonals of $\sigma^2\{\hat{Y}\} = \sigma^2 H$ and of $\sigma^2\{e\}$:

$$\text{Var}(\hat{Y}_i) = h_{ii} \sigma^2 \geq 0, \quad \text{Var}(e_i) = (1 - h_{ii}) \sigma^2 \geq 0.$$

Properties of the residuals

$e = (I - H)Y$ is again a linear function of Y . Since $(I - H)X = 0$:

$$E(e) = (I - H)X\beta = 0, \quad \sigma^2\{e\} = \sigma^2(I - H),$$

and with normal errors, $e \sim N(0, \sigma^2(I - H))$.

The proof of $\sigma^2\{e\}$ is Assignment 2, Question 3(c).

$H \cdot Y$

Write h_{ii} for the diagonal entries of H . On the diagonals of $\sigma^2\{\hat{Y}\} = \sigma^2 H$ and of $\sigma^2\{e\}$:

$$\text{Var}(\hat{Y}_i) = h_{ii} \sigma^2 \geq 0, \quad \text{Var}(e_i) = (1 - h_{ii}) \sigma^2 \geq 0.$$

Hence $0 \leq h_{ii} \leq 1$ for every i .

Fitted values and residuals share the variance

Adding the two variance-covariance matrices:

$$\sigma^2\{\hat{Y}\} + \sigma^2\{e\} = \sigma^2\mathbf{H} + \sigma^2(\mathbf{I} - \mathbf{H}) = \sigma^2\mathbf{I} = \sigma^2\{Y\}.$$

Fitted values and residuals share the variance

Adding the two variance-covariance matrices:

$$\sigma^2\{\hat{Y}\} + \sigma^2\{e\} = \sigma^2\mathbf{H} + \sigma^2(\mathbf{I} - \mathbf{H}) = \sigma^2\mathbf{I} = \sigma^2\{Y\}.$$

For each observation:

$$\underbrace{Var(Y_i)}_{\sigma^2} = \underbrace{Var(\hat{Y}_i)}_{h_{ii} \sigma^2} + \underbrace{Var(e_i)}_{(1-h_{ii}) \sigma^2}.$$

Fitted values and residuals share the variance

Adding the two variance-covariance matrices:

$$\sigma^2\{\hat{\mathbf{Y}}\} + \sigma^2\{\mathbf{e}\} = \sigma^2\mathbf{H} + \sigma^2(\mathbf{I} - \mathbf{H}) = \sigma^2\mathbf{I} = \sigma^2\{\mathbf{Y}\}.$$

For each observation:

$$\underbrace{\text{Var}(Y_i)}_{\sigma^2} = \underbrace{\text{Var}(\hat{Y}_i)}_{h_{ii} \sigma^2} + \underbrace{\text{Var}(e_i)}_{(1-h_{ii}) \sigma^2}.$$

Since $\hat{Y}_i = \sum_j h_{ij} Y_j$, h_{ii} is the weight of Y_i in its own fitted value.

- With a large h_{ii} , the fit is pulled towards Y_i : $\hat{Y}_i$ varies a lot and e_i little.
- With a small h_{ii} , the opposite.

Fitted values and residuals share the variance

Adding the two variance-covariance matrices:

$$\sigma^2\{\hat{\mathbf{Y}}\} + \sigma^2\{\mathbf{e}\} = \sigma^2\mathbf{H} + \sigma^2(\mathbf{I} - \mathbf{H}) = \sigma^2\mathbf{I} = \sigma^2\{\mathbf{Y}\}.$$

For each observation:

$$\underbrace{\text{Var}(Y_i)}_{\sigma^2} = \underbrace{\text{Var}(\hat{Y}_i)}_{h_{ii} \sigma^2} + \underbrace{\text{Var}(e_i)}_{(1-h_{ii}) \sigma^2}.$$

Since $\hat{Y}_i = \sum_j h_{ij} Y_j$, h_{ii} is the weight of Y_i in its own fitted value.

- With a large h_{ii} , the fit is pulled towards Y_i : $\hat{Y}_i$ varies a lot and e_i little.
- With a small h_{ii} , the opposite.

The residuals do not all have the same variance, even when the errors do.

$$\text{so } \text{Var}(e_i) \leq \text{Var}(e_i) = \sigma^2$$

Exercise: a model with only an intercept

$$H = X(X^T X)^{-1} X^T; \quad X^T X = (1 \dots 1) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = n; \quad H = n^{-1} X X^T \\ = n^{-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} (1 \dots 1)$$

Consider the model

$$Y_i = \beta_0 + \epsilon_i, \quad i = 1, \dots, n.$$

1. What is the design matrix $\mathbf{X}$?

$$X = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

2. What is the value of h_{ii} , for any i ?

$$h_{ii} = 1/n$$

3. What is the variance of $\hat{Y}_i$? $\text{Var}(\hat{Y}_i) = h_{ii} \sigma^2 = \sigma^2/n$

4. This is a well-known formula for the variance of an estimator. Which one?

5. Is this result surprising? Explain.

$$\hat{Y}_i = \hat{\beta}_0 \Rightarrow \text{Var}(\hat{Y}_i) = \text{Var}(\hat{\beta}_0)$$

Exercise: solution

1. $\mathbf{X} = \mathbf{1}$, the column of n ones, so $\mathbf{X}'\mathbf{X} = \mathbf{1}'\mathbf{1} = n$ and $(\mathbf{X}'\mathbf{X})^{-1} = 1/n$.

Exercise: solution

1. $\mathbf{X} = \mathbf{1}$, the column of n ones, so $\mathbf{X}'\mathbf{X} = \mathbf{1}'\mathbf{1} = n$ and $(\mathbf{X}'\mathbf{X})^{-1} = 1/n$.
2. $\mathbf{H} = \mathbf{1} \frac{1}{n} \mathbf{1}' = \frac{1}{n} \mathbf{J}$, where $\mathbf{J} = \mathbf{1}\mathbf{1}'$ is the $n \times n$ matrix of ones. Every entry of $\mathbf{H}$ is $1/n$: $h_{ii} = 1/n$.

Exercise: solution

1. $\mathbf{X} = \mathbf{1}$, the column of n ones, so $\mathbf{X}'\mathbf{X} = \mathbf{1}'\mathbf{1} = n$ and $(\mathbf{X}'\mathbf{X})^{-1} = 1/n$.
2. $\mathbf{H} = \mathbf{1} \frac{1}{n} \mathbf{1}' = \frac{1}{n} \mathbf{J}$, where $\mathbf{J} = \mathbf{1}\mathbf{1}'$ is the $n \times n$ matrix of ones. Every entry of $\mathbf{H}$ is $1/n$: $h_{ii} = 1/n$.
3. $\text{Var}(\hat{Y}_i) = h_{ii} \sigma^2 = \sigma^2/n$.

Exercise: solution

1. $\mathbf{X} = \mathbf{1}$, the column of n ones, so $\mathbf{X}'\mathbf{X} = \mathbf{1}'\mathbf{1} = n$ and $(\mathbf{X}'\mathbf{X})^{-1} = 1/n$.
2. $\mathbf{H} = \mathbf{1} \frac{1}{n} \mathbf{1}' = \frac{1}{n} \mathbf{J}$, where $\mathbf{J} = \mathbf{1}\mathbf{1}'$ is the $n \times n$ matrix of ones. Every entry of $\mathbf{H}$ is $1/n$: $h_{ii} = 1/n$.
3. $\text{Var}(\hat{Y}_i) = h_{ii} \sigma^2 = \sigma^2/n$.
4. It is $\text{Var}(\bar{Y}) = \sigma^2/n$: indeed $\hat{\mathbf{Y}} = \frac{1}{n} \mathbf{J}\mathbf{Y} = \bar{Y}\mathbf{1}$, so $\hat{Y}_i = \bar{Y}$ for every i .

Exercise: solution

1. $\mathbf{X} = \mathbf{1}$, the column of n ones, so $\mathbf{X}'\mathbf{X} = \mathbf{1}'\mathbf{1} = n$ and $(\mathbf{X}'\mathbf{X})^{-1} = 1/n$.
2. $\mathbf{H} = \mathbf{1} \frac{1}{n} \mathbf{1}' = \frac{1}{n} \mathbf{J}$, where $\mathbf{J} = \mathbf{1}\mathbf{1}'$ is the $n \times n$ matrix of ones. Every entry of $\mathbf{H}$ is $1/n$: $h_{ii} = 1/n$.
3. $\text{Var}(\hat{Y}_i) = h_{ii} \sigma^2 = \sigma^2/n$.
4. It is $\text{Var}(\bar{Y}) = \sigma^2/n$: indeed $\hat{\mathbf{Y}} = \frac{1}{n} \mathbf{J}\mathbf{Y} = \bar{Y}\mathbf{1}$, so $\hat{Y}_i = \bar{Y}$ for every i .
5. Not surprising: least squares minimizes $\sum_i (Y_i - \beta_0)^2$, and setting the derivative $-2 \sum_i (Y_i - \beta_0)$ to zero gives $\hat{\beta}_0 = \bar{Y}$. With no predictor, every fitted value is the sample mean.

Pythagoras and the analysis of variance

Since $\mathbf{1}$ is a column of $\mathbf{X}$, the vector $\bar{Y}\mathbf{1}$ is in $C(\mathbf{X})$, and so is $\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}$. Then

$$\mathbf{Y} - \bar{Y}\mathbf{1} = (\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}) + \mathbf{e}, \quad \text{with } (\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}) \perp \mathbf{e}.$$

$$Y = \hat{Y} + e$$

↪ the first column
of X

Pythagoras and the analysis of variance

Since $\mathbf{1}$ is a column of $\mathbf{X}$, the vector $\bar{Y}\mathbf{1}$ is in $C(\mathbf{X})$, and so is $\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}$. Then

$$\mathbf{Y} - \bar{Y}\mathbf{1} = (\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}) + \mathbf{e}, \quad \text{with } (\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}) \perp \mathbf{e}.$$

$$= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad \text{and} \quad \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Pythagoras:

$$\underbrace{\|\mathbf{Y} - \bar{Y}\mathbf{1}\|^2}_{SST} = \underbrace{\|\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}\|^2}_{SSR} + \underbrace{\|\mathbf{e}\|^2}_{SSE}.$$

This is the decomposition of Lecture 6, now for any number of predictors.

$$\|\mathbf{y} - \bar{y}\mathbf{1}\|^2 = \left\| \begin{pmatrix} y_1 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix} \right\|^2 = \sum_{i=1}^n (y_i - \bar{y})^2$$

Pythagoras and the analysis of variance

Since $\mathbf{1}$ is a column of $\mathbf{X}$, the vector $\bar{Y}\mathbf{1}$ is in $C(\mathbf{X})$, and so is $\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}$. Then

$$\mathbf{Y} - \bar{Y}\mathbf{1} = (\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}) + \mathbf{e}, \quad \text{with } (\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}) \perp \mathbf{e}.$$

Pythagoras:

$$\underbrace{\|\mathbf{Y} - \bar{Y}\mathbf{1}\|^2}_{SST} = \underbrace{\|\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}\|^2}_{SSR} + \underbrace{\|\mathbf{e}\|^2}_{SSE}.$$

This is the decomposition of Lecture 6, now for any number of predictors.

As in Lecture 6, the **coefficient of determination**

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

is the proportion of the variation of Y explained by the regression, now with any number of predictors.

Degrees of freedom are dimensions

Each sum of squares is the squared length of a vector that lives in a subspace:

	Vector	lives in a space of dimension	degrees of freedom
$\in \mathbb{R}^m$	$\mathbf{Y} - \bar{Y}\mathbf{1}$	$n - 1$ (orthogonal to $\mathbf{1}$)	$n - 1$ for SST
$\in \mathbb{R}^p$	$\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}$	$p - 1$ ($C(\mathbf{X})$, orthogonal to $\mathbf{1}$)	$p - 1$ for SSR
$\in \mathbb{R}^{m-p}$	$\mathbf{e}$	$n - p$ (orthogonal to $C(\mathbf{X})$)	$n - p$ for SSE

Degrees of freedom are dimensions

Each sum of squares is the squared length of a vector that lives in a subspace:

Vector	lives in a space of dimension	degrees of freedom
$\mathbf{Y} - \bar{Y}\mathbf{1}$	$n - 1$ (orthogonal to $\mathbf{1}$)	$n - 1$ for SST
$\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}$	$p - 1$ ($C(\mathbf{X})$, orthogonal to $\mathbf{1}$)	$p - 1$ for SSR
$\mathbf{e}$	$n - p$ (orthogonal to $C(\mathbf{X})$)	$n - p$ for SSE

Why orthogonal to $\mathbf{1}$: with $\mathbf{J} = \mathbf{1}\mathbf{1}'$, $\bar{Y}\mathbf{1} = \frac{1}{n}\mathbf{J}\mathbf{Y}$, so $\mathbf{Y} - \bar{Y}\mathbf{1} = (\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{Y}$, and

$$\mathbf{1}'\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right) = \mathbf{1}' - \frac{1}{n}(\mathbf{1}'\mathbf{1})\mathbf{1}' = \mathbf{1}' - \mathbf{1}' = \mathbf{0}'.$$

Likewise for $\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}$, since $\mathbf{1}'\hat{\mathbf{Y}} = \mathbf{1}'\mathbf{Y}$ (because $\mathbf{1}'\mathbf{e} = 0$).

Degrees of freedom are dimensions

Each sum of squares is the squared length of a vector that lives in a subspace:

Vector	lives in a space of dimension	degrees of freedom
$\mathbf{Y} - \bar{Y}\mathbf{1}$	$n - 1$ (orthogonal to $\mathbf{1}$)	$n - 1$ for SST
$\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}$	$p - 1$ ($C(\mathbf{X})$, orthogonal to $\mathbf{1}$)	$p - 1$ for SSR
$\mathbf{e}$	$n - p$ (orthogonal to $C(\mathbf{X})$)	$n - p$ for SSE

Why orthogonal to $\mathbf{1}$: with $\mathbf{J} = \mathbf{1}\mathbf{1}'$, $\bar{Y}\mathbf{1} = \frac{1}{n}\mathbf{J}\mathbf{Y}$, so $\mathbf{Y} - \bar{Y}\mathbf{1} = (\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{Y}$, and

$$\mathbf{1}'\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right) = \mathbf{1}' - \frac{1}{n}(\mathbf{1}'\mathbf{1})\mathbf{1}' = \mathbf{1}' - \mathbf{1}' = \mathbf{0}'.$$

Likewise for $\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}$, since $\mathbf{1}'\hat{\mathbf{Y}} = \mathbf{1}'\mathbf{Y}$ (because $\mathbf{1}'\mathbf{e} = 0$).

The degrees of freedom add up as the dimensions do: $n - 1 = (p - 1) + (n - p)$.
With one predictor, $p = 2$: as in Lecture 6.

In R: the acceleration data

```
X <- model.matrix(fit2)
e <- resid(fit2); Yhat <- fitted(fit2)
round(c(t(X) %*% e, sum(Yhat * e)), 8)      # all zero

## [1] 0 0 0 0

Y <- autmpg$acc
c(SST = sum((Y - mean(Y))^2),
  SSR_plus_SSE = sum((Yhat - mean(Y))^2) + sum(e^2))

##          SST SSR_plus_SSE
## 2969.013    2969.013
```

In R: the acceleration data

```
X <- model.matrix(fit2)
e <- resid(fit2); Yhat <- fitted(fit2)
round(c(t(X) %*% e, sum(Yhat * e)), 8)      # all zero

## [1] 0 0 0 0

Y <- autmpg$acc
c(SST = sum((Y - mean(Y))^2),
  SSR_plus_SSE = sum((Yhat - mean(Y))^2) + sum(e^2))

##          SST SSR_plus_SSE
##    2969.013    2969.013
```

Next: tests and intervals for the coefficients. The sums of squares come back for the overall F test.

Inference on one coefficient

Confidence intervals and tests for β_k

With normal errors, $\hat{\beta} \sim N(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$, and $s^2\{\hat{\beta}_k\}$ is the diagonal entry k of $MSE(\mathbf{X}'\mathbf{X})^{-1}$.

Confidence intervals and tests for β_k

With normal errors, $\hat{\beta} \sim N(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$, and $s^2\{\hat{\beta}_k\}$ is the diagonal entry k of $MSE(\mathbf{X}'\mathbf{X})^{-1}$.

As in simple linear regression, with $n - p$ degrees of freedom instead of $n - 2$:

For β_k the true
value

$$\frac{\hat{\beta}_k - \beta_k}{s\{\hat{\beta}_k\}} \sim t(n - p).$$

Confidence intervals and tests for β_k

With normal errors, $\hat{\beta} \sim N(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$, and $s^2\{\hat{\beta}_k\}$ is the diagonal entry k of $MSE(\mathbf{X}'\mathbf{X})^{-1}$.

As in simple linear regression, with $n - p$ degrees of freedom instead of $n - 2$:

$$\frac{\hat{\beta}_k - \beta_k}{s\{\hat{\beta}_k\}} \sim t(n - p).$$

- Confidence interval: $\hat{\beta}_k \pm t(1 - \alpha/2; n - p) s\{\hat{\beta}_k\}$.
- Test of $H_0 : \beta_k = 0$ against $H_a : \beta_k \neq 0$: $t^* = \hat{\beta}_k / s\{\hat{\beta}_k\}$, compared with $t(1 - \alpha/2; n - p)$.

Reject exactly when 0 not in
the CI

Confidence intervals and tests for β_k

With normal errors, $\hat{\beta} \sim N(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$, and $s^2\{\hat{\beta}_k\}$ is the diagonal entry k of $MSE(\mathbf{X}'\mathbf{X})^{-1}$.

As in simple linear regression, with $n - p$ degrees of freedom instead of $n - 2$:

$$\frac{\hat{\beta}_k - \beta_k}{s\{\hat{\beta}_k\}} \sim t(n - p).$$

- Confidence interval: $\hat{\beta}_k \pm t(1 - \alpha/2; n - p) s\{\hat{\beta}_k\}$.
- Test of $H_0 : \beta_k = 0$ against $H_a : \beta_k \neq 0$: $t^* = \hat{\beta}_k / s\{\hat{\beta}_k\}$, compared with $t(1 - \alpha/2; n - p)$.

$H_0 : \beta_k = 0$ asks whether X_k adds anything once the other predictors are in the model.

The acceleration data

```
fit2 <- lm(acc ~ wt + hp, data = autmpg)
confint(fit2, level = 0.95)
```

```
##                2.5 %          97.5 %
## (Intercept) 17.827628208 19.111556618
## wt          0.001885979  0.002697749
## hp          -0.102212640 -0.084304701
```

The acceleration data

```
fit2 <- lm(acc ~ wt + hp, data = automp)  
confint(fit2, level = 0.95)
```

```
##                2.5 %          97.5 %  
## (Intercept) 17.827628208 19.111556618  
## wt          0.001885979  0.002697749  
## hp          -0.102212640 -0.084304701
```

Among cars with the same horsepower, 1000 lb more weight adds between 1.9 and 2.7 seconds to the mean time to 60 mph (95% confidence).

The acceleration data

```
fit2 <- lm(acc ~ wt + hp, data = automp)  
confint(fit2, level = 0.95)
```

```
##                2.5 %          97.5 %  
## (Intercept) 17.827628208 19.111556618  
## wt          0.001885979  0.002697749  
## hp          -0.102212640 -0.084304701
```

Among cars with the same horsepower, 1000 lb more weight adds between 1.9 and 2.7 seconds to the mean time to 60 mph (95% confidence).

Among cars of the same weight, 10 more horsepower take between 0.84 and 1.02 seconds off it.

The acceleration data

```
fit2 <- lm(acc ~ wt + hp, data = automp)  
confint(fit2, level = 0.95)
```

```
##                2.5 %          97.5 %  
## (Intercept) 17.827628208 19.111556618  
## wt          0.001885979  0.002697749  
## hp          -0.102212640 -0.084304701
```

Among cars with the same horsepower, 1000 lb more weight adds between 1.9 and 2.7 seconds to the mean time to 60 mph (95% confidence).

Among cars of the same weight, 10 more horsepower take between 0.84 and 1.02 seconds off it.

Neither interval contains 0: both predictors matter, each given the other.

Mean responses and new observations

Intervals at a new point $\mathbf{X}_h$

From Lecture ¹¹10, $\hat{Y}_h = \mathbf{X}_h' \hat{\boldsymbol{\beta}}$ with

$$s^2\{\hat{Y}_h\} = MSE \mathbf{X}_h' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_h, \quad s^2\{\text{pred}\} = MSE + s^2\{\hat{Y}_h\}.$$

Intervals at a new point $\mathbf{X}_h$

From Lecture 10, $\hat{Y}_h = \mathbf{X}_h' \hat{\boldsymbol{\beta}}$ with

$$s^2\{\hat{Y}_h\} = MSE \mathbf{X}_h' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_h, \quad s^2\{\text{pred}\} = MSE + s^2\{\hat{Y}_h\}.$$

- Confidence interval for the mean response $E(Y_h)$:
 $\hat{Y}_h \pm t(1 - \alpha/2; n - p) s\{\hat{Y}_h\}.$
- Prediction interval for a new observation $Y_{h(\text{new})}$:
 $\hat{Y}_h \pm t(1 - \alpha/2; n - p) s\{\text{pred}\}.$

Intervals at a new point $\mathbf{X}_h$

From Lecture 10, $\hat{Y}_h = \mathbf{X}_h' \hat{\beta}$ with

$$s^2\{\hat{Y}_h\} = MSE \mathbf{X}_h' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_h, \quad s^2\{\text{pred}\} = MSE + s^2\{\hat{Y}_h\}.$$

- Confidence interval for the mean response $E(Y_h)$:
 $\hat{Y}_h \pm t(1 - \alpha/2; n - p) s\{\hat{Y}_h\}.$
- Prediction interval for a new observation $Y_{h(\text{new})}$:
 $\hat{Y}_h \pm t(1 - \alpha/2; n - p) s\{\text{pred}\}.$

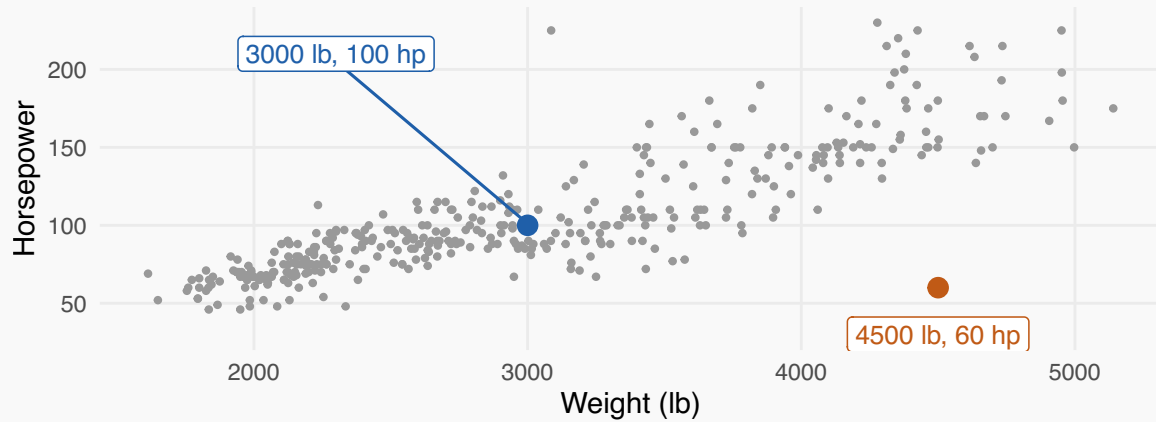
```
new_car <- data.frame(wt = 3000, hp = 100)
predict(fit2, new_car, interval = "confidence")
```

```
##           fit           lwr           upr
## 1 16.01932 15.83937 16.19927
```

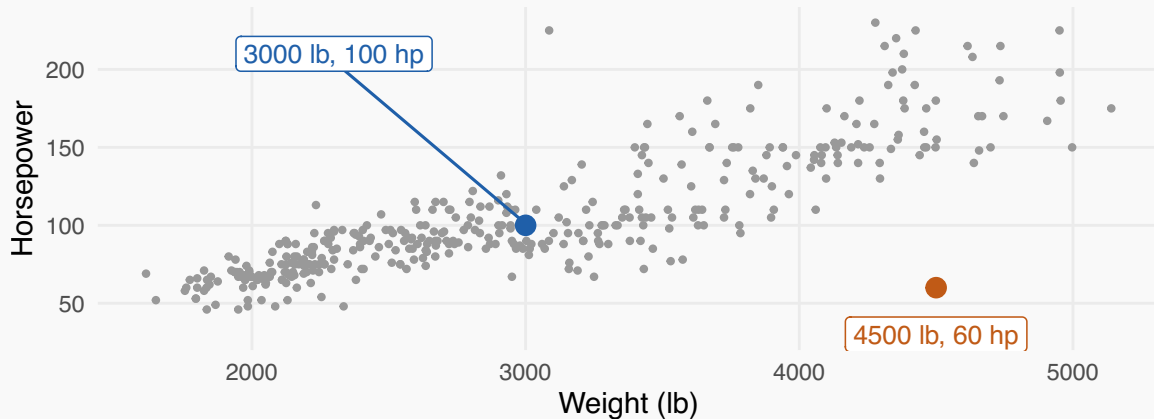
```
predict(fit2, new_car, interval = "prediction")
```

```
##           fit           lwr           upr
## 1 16.01932 12.58971 19.44892
```

Hidden extrapolation

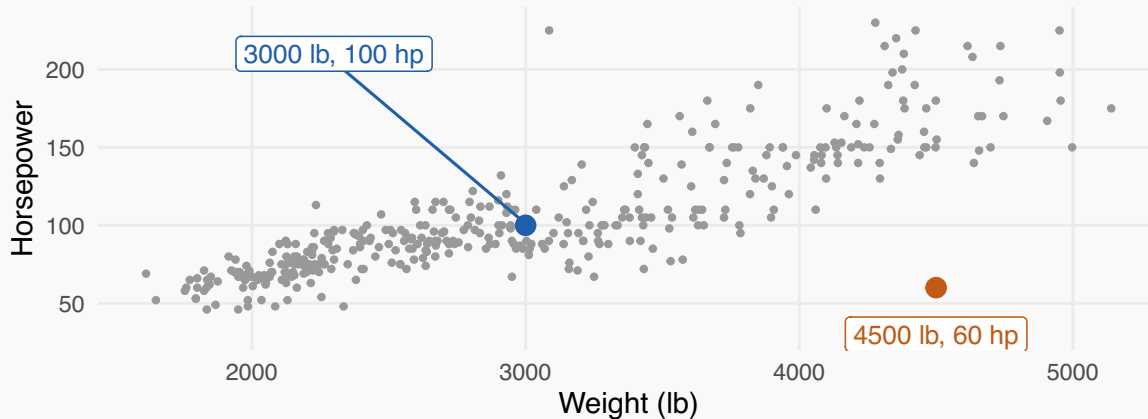


Hidden extrapolation



A car of 4500 lb and 60 hp is within the range of each predictor, but **outside the cloud of the data**: no heavy car in the data has a small engine.

Hidden extrapolation



A car of 4500 lb and 60 hp is within the range of each predictor, but **outside the cloud of the data**: no heavy car in the data has a small engine.

Its $\mathbf{X}'_h(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h$ is 0.085, against 0.0028 for the car of 3000 lb and 100 hp: its confidence interval is 5.5 times wider, and it rests on the model holding where there are no data.

Analysis of variance

Sums of squares as quadratic forms

$$A\mathbf{y} \rightarrow \mathbf{y}' A \mathbf{y}$$

$$SSE = \|\mathbf{y} - \hat{\boldsymbol{\beta}}\|^2 = (\mathbf{y} - H\mathbf{y})' (\mathbf{y} - H\mathbf{y})$$

Recall $\mathbf{J} = \mathbf{1}\mathbf{1}'$, the $n \times n$ matrix of ones, and $\bar{Y}\mathbf{1} = \frac{1}{n}\mathbf{J}\mathbf{Y}$. Then

$$SST = \mathbf{Y}'\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)\mathbf{Y}, \quad SSE = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y},$$

$$SSR = \mathbf{Y}'\left(\mathbf{H} - \frac{1}{n}\mathbf{J}\right)\mathbf{Y}.$$

Sums of squares as quadratic forms

Recall $\mathbf{J} = \mathbf{1}\mathbf{1}'$, the $n \times n$ matrix of ones, and $\bar{Y}\mathbf{1} = \frac{1}{n}\mathbf{J}\mathbf{Y}$. Then

$$SST = \mathbf{Y}'\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)\mathbf{Y}, \quad SSE = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y},$$

$$SSR = \mathbf{Y}'\left(\mathbf{H} - \frac{1}{n}\mathbf{J}\right)\mathbf{Y}.$$

Each is a **quadratic form** $\mathbf{Y}'\mathbf{A}\mathbf{Y} = \sum_i \sum_j a_{ij}Y_iY_j$, with $\mathbf{A}$ symmetric.

Sums of squares as quadratic forms

Recall $\mathbf{J} = \mathbf{1}\mathbf{1}'$, the $n \times n$ matrix of ones, and $\bar{Y}\mathbf{1} = \frac{1}{n}\mathbf{J}\mathbf{Y}$. Then

$$\underline{SST} = \mathbf{Y}'\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)\mathbf{Y}, \quad \underline{SSE} = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y},$$

$$\underline{SSR} = \mathbf{Y}'\left(\mathbf{H} - \frac{1}{n}\mathbf{J}\right)\mathbf{Y}.$$

Each is a **quadratic form** $\mathbf{Y}'\mathbf{A}\mathbf{Y} = \sum_i \sum_j a_{ij}Y_iY_j$, with $\mathbf{A}$ symmetric.

Each matrix $\mathbf{A}$ here is a projection: symmetric and idempotent.

So $\mathbf{Y}'\mathbf{A}\mathbf{Y} = (\mathbf{A}\mathbf{Y})'(\mathbf{A}\mathbf{Y}) = \|\mathbf{A}\mathbf{Y}\|^2$: the squared length of a projection, as in Lecture 11.

Why these formulas hold

If A is symmetric and idempotent: $\|AY\|^2 = Y' A' A Y = Y' A A Y = Y' A Y$.

Why these formulas hold

If A is symmetric and idempotent: $\|AY\|^2 = Y'A'AY = Y'AA Y = Y'AY$.

$SSE = \|e\|^2$ with $e = (I - H)Y$, and $I - H$ is symmetric and idempotent (Lecture 11), so $SSE = Y'(I - H)Y$.

Why these formulas hold

If $\mathbf{A}$ is symmetric and idempotent: $\|\mathbf{A}\mathbf{Y}\|^2 = \mathbf{Y}'\mathbf{A}'\mathbf{A}\mathbf{Y} = \mathbf{Y}'\mathbf{A}\mathbf{A}\mathbf{Y} = \mathbf{Y}'\mathbf{A}\mathbf{Y}$.

$SSE = \|\mathbf{e}\|^2$ with $\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$, and $\mathbf{I} - \mathbf{H}$ is symmetric and idempotent (Lecture 11), so $SSE = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$.

$SST = \|\mathbf{Y} - \bar{Y}\mathbf{1}\|^2$ with $\mathbf{Y} - \bar{Y}\mathbf{1} = (\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{Y}$. This matrix is symmetric, and idempotent since $\mathbf{J}\mathbf{J} = \mathbf{1}(\mathbf{1}'\mathbf{1})\mathbf{1}' = n\mathbf{J}$:

$$\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right) = \mathbf{I} - \frac{2}{n}\mathbf{J} + \frac{1}{n^2}n\mathbf{J} = \mathbf{I} - \frac{1}{n}\mathbf{J}.$$

Why these formulas hold

If $\mathbf{A}$ is symmetric and idempotent: $\|\mathbf{A}\mathbf{Y}\|^2 = \mathbf{Y}'\mathbf{A}'\mathbf{A}\mathbf{Y} = \mathbf{Y}'\mathbf{A}\mathbf{A}\mathbf{Y} = \mathbf{Y}'\mathbf{A}\mathbf{Y}$.

$SSE = \|\mathbf{e}\|^2$ with $\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$, and $\mathbf{I} - \mathbf{H}$ is symmetric and idempotent (Lecture 11), so $SSE = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$.

$SST = \|\mathbf{Y} - \bar{Y}\mathbf{1}\|^2$ with $\mathbf{Y} - \bar{Y}\mathbf{1} = (\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{Y}$. This matrix is symmetric, and idempotent since $\mathbf{J}\mathbf{J} = \mathbf{1}(\mathbf{1}'\mathbf{1})\mathbf{1}' = n\mathbf{J}$:

$$\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right) = \mathbf{I} - \frac{2}{n}\mathbf{J} + \frac{1}{n^2}n\mathbf{J} = \mathbf{I} - \frac{1}{n}\mathbf{J}.$$

$SSR = SST - SSE$, by Pythagoras (slide 7):

$$SSR = \mathbf{Y}'\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)\mathbf{Y} - \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y} = \mathbf{Y}'\left(\mathbf{H} - \frac{1}{n}\mathbf{J}\right)\mathbf{Y}.$$

$\mathbf{H} - \frac{1}{n}\mathbf{J}$ is symmetric and idempotent too: $\mathbf{H}\mathbf{1} = \mathbf{1}$ (as $\mathbf{1}$ is a column of $\mathbf{X}$) gives $\mathbf{H}\mathbf{J} = \mathbf{J}\mathbf{H} = \mathbf{J}$.

The mean of a quadratic form

$$Y^T A Y \quad \text{with} \quad \mathbb{E}[Y] = \mu, \quad \sigma^2\{Y\} = \Sigma$$

Trace: $\text{tr}(\mathbf{A}) = \sum_i a_{ii}$, and $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$.

The mean of a quadratic form

Trace: $tr(\mathbf{A}) = \sum_i a_{ii}$, and $tr(\mathbf{AB}) = tr(\mathbf{BA})$.

For a random vector $\mathbf{Y}$ with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$,

$$E(\mathbf{Y}'\mathbf{A}\mathbf{Y}) = tr(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}.$$

The mean of a quadratic form

Trace: $tr(\mathbf{A}) = \sum_i a_{ii}$, and $tr(\mathbf{AB}) = tr(\mathbf{BA})$.

For a random vector $\mathbf{Y}$ with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$,

$$E(\mathbf{Y}'\mathbf{A}\mathbf{Y}) = tr(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}.$$

Why. A number is its own trace, so $\mathbf{Y}'\mathbf{A}\mathbf{Y} = tr(\mathbf{A}\mathbf{Y}\mathbf{Y}')$, and

$$E(\mathbf{Y}'\mathbf{A}\mathbf{Y}) = tr(\mathbf{A} E(\mathbf{Y}\mathbf{Y}')) = tr(\mathbf{A}(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}')) = tr(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}.$$

The mean of a quadratic form

Trace: $tr(\mathbf{A}) = \sum_i a_{ii}$, and $tr(\mathbf{AB}) = tr(\mathbf{BA})$.

For a random vector $\mathbf{Y}$ with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$,

$$E(\mathbf{Y}'\mathbf{A}\mathbf{Y}) = tr(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}.$$

Why. A number is its own trace, so $\mathbf{Y}'\mathbf{A}\mathbf{Y} = tr(\mathbf{A}\mathbf{Y}\mathbf{Y}')$, and

$$E(\mathbf{Y}'\mathbf{A}\mathbf{Y}) = tr(\mathbf{A} E(\mathbf{Y}\mathbf{Y}')) = tr(\mathbf{A}(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}')) = tr(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}.$$

The analogue of $E(Y^2) = Var(Y) + E(Y)^2$.

MSE is unbiased

Apply it to $SSE = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$, with $\Sigma = \sigma^2 \mathbf{I}$ and $\mu = \mathbf{X}\beta$.

MSE is unbiased

$$Y \rightarrow E[Y] = X\beta = \mu$$
$$\rightarrow \sigma^2 \{Y\} = \sigma^2 I = \Sigma$$

Apply it to $SSE = Y'(I - H)Y$, with $\Sigma = \sigma^2 I$ and $\mu = X\beta$.

- $tr(H) = tr(X(X'X)^{-1}X') = tr((X'X)^{-1}X'X) = tr(I_p) = p$, so $tr((I - H)\sigma^2 I) = (n - p)\sigma^2$.
- $(I - H)X = 0$, so $\mu'(I - H)\mu = 0$.

MSE is unbiased

$$MSE = \frac{SSE}{n-p}$$

Apply it to $SSE = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$, with $\Sigma = \sigma^2 \mathbf{I}$ and $\mu = \mathbf{X}\beta$.

- $tr(\mathbf{H}) = tr(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = tr((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}) = tr(\mathbf{I}_p) = p$, so $tr((\mathbf{I} - \mathbf{H})\sigma^2 \mathbf{I}) = (n - p)\sigma^2$.
- $(\mathbf{I} - \mathbf{H})\mathbf{X} = \mathbf{0}$, so $\mu'(\mathbf{I} - \mathbf{H})\mu = 0$.

Hence

$$E(SSE) = (n - p) \sigma^2, \quad E(MSE) = E\left(\frac{SSE}{n - p}\right) = \sigma^2.$$

MSE is unbiased

Apply it to $SSE = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$, with $\Sigma = \sigma^2 \mathbf{I}$ and $\mu = \mathbf{X}\beta$.

- $tr(\mathbf{H}) = tr(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = tr((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}) = tr(\mathbf{I}_p) = p$, so $tr((\mathbf{I} - \mathbf{H})\sigma^2 \mathbf{I}) = (n - p)\sigma^2$.
- $(\mathbf{I} - \mathbf{H})\mathbf{X} = \mathbf{0}$, so $\mu'(\mathbf{I} - \mathbf{H})\mu = 0$.

Hence

$$E(SSE) = (n - p) \sigma^2, \quad E(MSE) = E\left(\frac{SSE}{n - p}\right) = \sigma^2.$$

The trace of a projection is the dimension of the space it projects on: $n - p$, again.

The mean of MSR

For $SSR = \mathbf{Y}'(\mathbf{H} - \frac{1}{n}\mathbf{J})\mathbf{Y}$: $tr(\mathbf{H} - \frac{1}{n}\mathbf{J}) = p - 1$, and since $\mathbf{H}\boldsymbol{\mu} = \boldsymbol{\mu}$,

$$\boldsymbol{\mu}'\left(\mathbf{H} - \frac{1}{n}\mathbf{J}\right)\boldsymbol{\mu} = \boldsymbol{\mu}'\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)\boldsymbol{\mu} = \sum_{i=1}^n (\mu_i - \bar{\mu})^2.$$

The mean of MSR

For $SSR = \mathbf{Y}'(\mathbf{H} - \frac{1}{n}\mathbf{J})\mathbf{Y}$: $tr(\mathbf{H} - \frac{1}{n}\mathbf{J}) = p - 1$, and since $\mathbf{H}\boldsymbol{\mu} = \boldsymbol{\mu}$,

$$\boldsymbol{\mu}'\left(\mathbf{H} - \frac{1}{n}\mathbf{J}\right)\boldsymbol{\mu} = \boldsymbol{\mu}'\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)\boldsymbol{\mu} = \sum_{i=1}^n (\mu_i - \bar{\mu})^2.$$

Hence, with $MSR = SSR/(p - 1)$,

$$E(MSR) = \sigma^2 + \frac{1}{p-1} \sum_{i=1}^n (\mu_i - \bar{\mu})^2 \geq \sigma^2 = E[MSE]$$

Handwritten notes above the equation:
 $x_i^T \beta$ (above μ_i)
 $\frac{1}{n} \sum_{i=1}^n x_i^T \beta$ (above $\bar{\mu}$)

So $\frac{E[MSR]}{E[MSE]} \geq 1$ with
 $\Leftrightarrow$ if and
 only if $\beta = 0$.

$\geq 0 \Leftrightarrow$ all $\mu_i = x_i^T \beta$
 are equal.

$\Leftrightarrow \beta = 0$

The mean of MSR

For $SSR = \mathbf{Y}'(\mathbf{H} - \frac{1}{n}\mathbf{J})\mathbf{Y}$: $tr(\mathbf{H} - \frac{1}{n}\mathbf{J}) = p - 1$, and since $\mathbf{H}\boldsymbol{\mu} = \boldsymbol{\mu}$,

$$\boldsymbol{\mu}'\left(\mathbf{H} - \frac{1}{n}\mathbf{J}\right)\boldsymbol{\mu} = \boldsymbol{\mu}'\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)\boldsymbol{\mu} = \sum_{i=1}^n (\mu_i - \bar{\mu})^2.$$

Hence, with $MSR = SSR/(p - 1)$,

$$E(MSR) = \sigma^2 + \frac{1}{p-1} \sum_{i=1}^n (\mu_i - \bar{\mu})^2 \geq \sigma^2.$$

Equality holds exactly when all the means $\mu_i = \beta_0 + \beta_1 X_{i1} + \cdots$ are equal, that is, when $\beta_1 = \cdots = \beta_{p-1} = 0$.

The mean of MSR

For $SSR = \mathbf{Y}'(\mathbf{H} - \frac{1}{n}\mathbf{J})\mathbf{Y}$: $tr(\mathbf{H} - \frac{1}{n}\mathbf{J}) = p - 1$, and since $\mathbf{H}\boldsymbol{\mu} = \boldsymbol{\mu}$,

$$\boldsymbol{\mu}'\left(\mathbf{H} - \frac{1}{n}\mathbf{J}\right)\boldsymbol{\mu} = \boldsymbol{\mu}'\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)\boldsymbol{\mu} = \sum_{i=1}^n (\mu_i - \bar{\mu})^2.$$

Hence, with $MSR = SSR/(p - 1)$,

$$E(MSR) = \sigma^2 + \frac{1}{p-1} \sum_{i=1}^n (\mu_i - \bar{\mu})^2 \geq \sigma^2.$$

Equality holds exactly when all the means $\mu_i = \beta_0 + \beta_1 X_{i1} + \cdots$ are equal, that is, when $\beta_1 = \cdots = \beta_{p-1} = 0$.

Large values of MSR/MSE are evidence that at least one predictor matters.

The ANOVA table

Source	SS	df	MS	F^*
Regression	SSR	$p - 1$	$MSR = SSR/(p - 1)$	MSR/MSE
Error	SSE	$n - p$	$MSE = SSE/(n - p)$	
Total	SST	$n - 1$		

The ANOVA table

Source	SS	df	MS	F^*
Regression	SSR	$p - 1$	$MSR = SSR/(p - 1)$	MSR/MSE
Error	SSE	$n - p$	$MSE = SSE/(n - p)$	
Total	SST	$n - 1$		

With one predictor, $p = 2$, this is the table of Lecture 6.

The overall F test

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$ against $H_a : \text{at least one } \beta_k \neq 0.$

The overall F test

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0 \quad \text{against} \quad H_a : \text{at least one } \beta_k \neq 0.$$

With normal errors, under H_0 , SSR/σ^2 and SSE/σ^2 are independent χ^2 variables with $p - 1$ and $n - p$ degrees of freedom (Cochran's theorem, which we take for granted), so

$$F^* = \frac{MSR}{MSE} \sim F(p - 1, n - p).$$

The overall F test

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0 \quad \text{against} \quad H_a : \text{at least one } \beta_k \neq 0.$$

With normal errors, under H_0 , SSR/σ^2 and SSE/σ^2 are independent χ^2 variables with $p - 1$ and $n - p$ degrees of freedom (Cochran's theorem, which we take for granted), so

$$F^* = \frac{MSR}{MSE} \sim F(p - 1, n - p).$$

Reject H_0 at level α when $F^* > F(1 - \alpha; p - 1, n - p)$.

The overall F test

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0 \quad \text{against} \quad H_a : \text{at least one } \beta_k \neq 0.$$

With normal errors, under H_0 , SSR/σ^2 and SSE/σ^2 are independent χ^2 variables with $p - 1$ and $n - p$ degrees of freedom (Cochran's theorem, which we take for granted), so

$$F^* = \frac{MSR}{MSE} \sim F(p - 1, n - p).$$

Reject H_0 at level α when $F^* > F(1 - \alpha; p - 1, n - p)$.

 Rejecting says that the predictors, *taken together*, explain some variation. It does not say which ones, nor that the model is good.

The acceleration data

Source	SS	df	MS	F^*
Regression	1794.7	2	897.3	295.7
Error	1174.3	387	3.034	
Total	2969.0	389		

The acceleration data

Source	SS	df	MS	F^*
Regression	1794.7	2	897.3	295.7
Error	1174.3	387	3.034	
Total	2969.0	389		

$F^* = 295.7 > F(0.95; 2, 387) = 3.02$: we reject H_0 . Weight and horsepower together are related to acceleration.

The acceleration data

Source	SS	df	MS	F^*
Regression	1794.7	2	897.3	295.7
Error	1174.3	387	3.034	
Total	2969.0	389		

$F^* = 295.7 > F(0.95; 2, 387) = 3.02$: we reject H_0 . Weight and horsepower together are related to acceleration.

In R, compare the model with the model without predictors: `anova(lm(acc ~ 1, data = autmpg), fit2)`.

The acceleration data

Source	SS	df	MS	F^*
Regression	1794.7	2	897.3	295.7
Error	1174.3	387	3.034	
Total	2969.0	389		

$F^* = 295.7 > F(0.95; 2, 387) = 3.02$: we reject H_0 . Weight and horsepower together are related to acceleration.

In R, compare the model with the model without predictors: `anova(lm(acc ~ 1, data = autmpg), fit2)`.

`anova(fit2)` alone gives a different table, split predictor by predictor: that is next week's extra sums of squares.

R^2 and adjusted R^2

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 0.604 \text{ here,} \quad R^2 = (\text{corr}(\mathbf{Y}, \hat{\mathbf{Y}}))^2.$$

R^2 and adjusted R^2

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 0.604 \text{ here,} \quad R^2 = (\text{corr}(\mathbf{Y}, \hat{\mathbf{Y}}))^2.$$

Adding a predictor can only lower SSE , so R^2 never decreases, even for a useless predictor. Adding a column of pure noise:

```
c(summary(fit2)$r.squared, summary(fit3)$r.squared)
## [1] 0.6044754 0.6045153
```

R^2 and adjusted R^2

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 0.604 \text{ here,} \quad R^2 = (\text{corr}(\mathbf{Y}, \hat{\mathbf{Y}}))^2.$$

Adding a predictor can only lower SSE , so R^2 never decreases, even for a useless predictor. Adding a column of pure noise:

```
c(summary(fit2)$r.squared, summary(fit3)$r.squared)
```

```
## [1] 0.6044754 0.6045153
```

The **adjusted** R^2 divides each sum of squares by its degrees of freedom:

$$R_a^2 = 1 - \frac{SSE/(n-p)}{SST/(n-1)} = 1 - \frac{n-1}{n-p} \frac{SSE}{SST},$$

and can decrease: here from 0.6024 to 0.6014 with the noise.

Dividing the numerator and the denominator of F^* by SST :

$$F^* = \frac{SSR/(p-1)}{SSE/(n-p)} = \frac{R^2/(p-1)}{(1-R^2)/(n-p)}.$$

Dividing the numerator and the denominator of F^* by SST :

$$F^* = \frac{SSR/(p-1)}{SSE/(n-p)} = \frac{R^2/(p-1)}{(1-R^2)/(n-p)}.$$

Here: $\frac{0.6045/2}{(1-0.6045)/387} = 295.7$, the F statistic of the table.

Dividing the numerator and the denominator of F^* by SST :

$$F^* = \frac{SSR/(p-1)}{SSE/(n-p)} = \frac{R^2/(p-1)}{(1-R^2)/(n-p)}.$$

Here: $\frac{0.6045/2}{(1-0.6045)/387} = 295.7$, the F statistic of the table.

For a fixed R^2 , F^* grows with n : with enough data, a weak relation is “significant”.

Reading summary(fit2)

```
s <- summary(fit2)
round(s$coefficients, 5)
```

##		Estimate	Std. Error	t value	Pr(> t)
##	(Intercept)	18.46959	0.32651	56.56593	0
##	wt	0.00229	0.00021	11.10182	0
##	hp	-0.09326	0.00455	-20.47774	0

```
round(c(sigma = s$sigma, R2 = s$r.squared,
        adj_R2 = s$adj.r.squared), 4)
```

##	sigma	R2	adj_R2
##	1.7420	0.6045	0.6024

Reading summary(fit2)

```
s <- summary(fit2)
round(s$coefficients, 5)
```

```
##              Estimate Std. Error   t value Pr(>|t|)
## (Intercept) 18.46959      0.32651  56.56593      0
## wt           0.00229      0.00021  11.10182      0
## hp          -0.09326      0.00455 -20.47774      0
```

```
round(c(sigma = s$sigma, R2 = s$r.squared,
        adj_R2 = s$adj.r.squared), 4)
```

```
##  sigma      R2 adj_R2
## 1.7420 0.6045 0.6024
```

Each line: $\hat{\beta}_k$, $s\{\hat{\beta}_k\}$, t^* on $n - p$ df and its p -value. Then $\sqrt{MSE}$, R^2 and R_a^2 ; the last line of `summary(fit2)` is the overall F test.

Reading summary(fit2)

```
s <- summary(fit2)
round(s$coefficients, 5)
```

```
##              Estimate Std. Error   t value Pr(>|t|)
## (Intercept) 18.46959      0.32651  56.56593      0
## wt           0.00229      0.00021  11.10182      0
## hp          -0.09326      0.00455 -20.47774      0
```

```
round(c(sigma = s$sigma, R2 = s$r.squared,
        adj_R2 = s$adj.r.squared), 4)
```

```
##  sigma      R2 adj_R2
## 1.7420 0.6045 0.6024
```

Each line: $\hat{\beta}_k$, $s\{\hat{\beta}_k\}$, t^* on $n - p$ df and its p -value. Then $\sqrt{MSE}$, R^2 and R_a^2 ; the last line of `summary(fit2)` is the overall F test.

Next week: testing several coefficients at once, extra sums of squares, multicollinearity, qualitative predictors.